Running Head: GENERALIZING RECOGNITION

Generalizing Recognition from Familiar to Novel Views

WONG Chun Nang

The Chinese University of Hong Kong

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Philosophy in Psychology

June 2001

Abstract

It has been proposed that objects are stored in view-based representations in the human visual system, i.e., information that we store about an object is tied to specific views of the object only. One problem about this view-based theory of visual object recognition concerns how view invariance is achieved through experience with a small number of views. The pooled activation account suggests the presence of view-specific representations, which are the most sensitive to their preferred view and also react to a range of neighboring views. Novel views activate different view-specific representations to different degrees, and the pooled activation determines recognition performance. Such an account predicts that when two views rather than one view have been studied, views between but not outside tend to activate the representations of the studied views to a larger extent, leading to interpolation but not extrapolation. It also predicts more evident interpolation when depth rotation does not occlude important parts of the objects than when occluded rotation occurs, because the view-specific representations of objects with non-occluded rotation are sensitive to a wider range of views and thus have more overlap with other representations. To test this prediction 125 university students, across four experiments, studied one or two views of a target object, and then made a same/different decision for a test image, which showed either the target or a distractor. Amoeboid and geon objects with and without self-occlusion during depth rotation were adopted. Results showed a general facilitation in recognition performance for test views that fall between the two studied views, although not in all conditions. With a large angular disparity between the two studied views, such internal facilitation was only evident if both studied views showed the same object features (i.e., there was no self-occlusion).

Having studied two very similar views resulted in subsequent recognition performance similar to the condition when only one of the views had been studied. Results supported and extended accounts of object recognition that rely on pooled activation mechanisms.

撮要

物體在人類的視覺系統中以影像爲本的表徵(view-based representations)儲存，即我們儲存的關於物體的資料是限於該物體於某角度顯示的影像。這個物體辨認的影像爲本理論要考慮的其中一個問題，是我們如何只從少數角度觀察過一物件，之後便能迅速辨認該物件而不受角度影響。聯合活動理論(pooled activation account)提出影像特定表徵(view-specific representation)之存在；它們對於物件於某一角度之影像最爲敏感，但同時也對鄰近的影像作出反應。一個全新的影像對不同的影像特定表徵作出不同程度的活化，而它們的聯合活動決定了物體辨認的表現。 根據這個理論，當兩個而非一個影像曾被經驗過，在它們中間的全新影像會較在外面的影像，對各個影像表徵產生較大的整體活化，導致內泛(interpolation)而非外泛(extrapolation)。這理論也預測物體部份在不同角度下若沒有互相遮蓋，則內泛的情況更顯著。這是因爲這類物體的影像特定表徵對較大範圍的影像敏感，因而互相有較多的重疊。爲著測試這些預測，一百二十五名大學學生參與了一個順序配對(sequential matching)的實驗。實驗利用了阿米巴狀及以 geon 構成的物體。結果顯示沒有自我遮蓋的物體有較多內泛，而外泛並不存在。此結果和聯合活動理論的說法吻合。

Acknowledgement

I would like to express my gratitude to Prof. William G. Hayward, my thesis supervisor, for his inspiring supervision throughout the whole course of this research project. I would also like to thank Prof. Trevor B. Penney, Prof. T.-C. Chan, Prof. Michael J. Tarr, and Prof. Isabel Gauthier for their comments on this report. Thanks also to Rebecca Lau Wing-Man, Wincy Lee Wing-Yee, Rexx Wong Hon-Chung, and Crystal Yuen-Wing Sze who helped to run the experiments.

View Generalization in Object Recognition by Pooled Activation

A fundamental issue in the field of human object recognition is how the human visual system achieves viewpoint-invariant recognition by combining information from a small number of object views. The present study focused on this problem, particularly for recognition at the subordinate level, i.e., discrimination of objects within a homogeneous object class. The main hypothesis tested here was that, when two views are studied, novel views within the range spanned by the experienced views (i.e., internal views) would be better recognized than novel views outside the range (i.e., external views). This prediction is based on the pooled activation account of view combination. The effects of self-occlusion and viewpoint difference between the studied views on this internal facilitation were also investigated.

Viewpoint-Sensitive Object Recognition at the Subordinate Level

One main problem in object recognition study is how our visual system is capable of recognizing an object despite constant changes in its properties such as position, scale, and orientation. Generally, theories proposed for human object recognition differ in what kinds of representations are formed and how they are formed. Viewpoint-insensitive theories suggest that representations are less tied to specific views and are thus generally unaffected by changes in different viewing conditions. Biederman's (1987, Biederman & Gerhardstein, 1993) recognition-by-components (RBC) model is representative of this type of theory. He proposed that objects can be divided into 3-D parts (geons) recoverable by simple 2-D features like edges and vertices. Therefore, as long as the geons and the spatial relationships between them remain visible, the object can be specified and recognized with virtually no difference in processing time across depth rotation, even if only one view has been experienced (Biederman & Bar, 1999; Biederman & Gerhardstein, 1995).

Inevitably, if information from one-shot exposure to an object were enough for its recognition at novel views, as proposed by the RBC model, the present research question would be meaningless. Nevertheless, numerous studies showed that our recognition system, at least for subordinate-level discrimination, is view-based, i.e., information that we store for an object is tied to specific views of the object only. Different information is available from different views of an object, and recognition performance is thus viewpoint sensitive. Behavioral studies have demonstrated, across a wide range of objects and tasks, that human recognition performance deteriorates systematically as the view of an object to be recognized deviates from its familiar view (Hayward & Tarr, 1997; Lawson, 1999; Lawson, Humphreys, & Watson, 1994; Liu, 1996; Newell & Findlay, 1997; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989; Tarr, Williams, Hayward, & Gauthier, 1998). This leads to the second class of theories, called view-based theories. View-based theories maintain that an object representation is a collection of view-dependent descriptions rather than a single or small number of descriptions of the view-insensitive structures.

One criticism of the view-based theories is that, at the first glance, an unlimited number of views have to be stored in memory to allow viewpoint-invariant recognition. However, studies in patients with brain lesions and research of the primate visual system suggested that this is not necessary. Weiskrantz & Saunders (1984), for example, trained monkeys to recognize objects at particular orientations or different patterns of illuminations. After a lesion of the anterior inferotemporal area (AIT), a general recognition deficiency occurred. Lesion of the posterior IT (PIT) and some prestriate areas, however, led to deterioration of recognition for transformed (in terms of size, orientation and illumination) images only. Recognition of original views remained intact. This finding suggests that we do not form view-invariant representations of objects in the first place.

Single cell recording of primate IT cortex showed view-tuned cells activated maximally by specific object views. Logothetis and Pauls (1995) trained three juvenile rhesus monkeys to recognize novel objects from one or two viewpoints. They then measured the behavioral performance of the monkeys in later recognition of the same objects at both familiar and novel viewpoints, and the concurrent IT neuron responses. Out of the 773 cells analyzed, they found 71 cells which reacted the most towards a certain trained object at a particular view, and systematically less rigorously as the object was rotated in depth. Only five out of the 773 neurons were equally active to all views of objects. Behaviorally, it was found that when only one object view was studied, later recognition performance dropped to about chance level when the test view was about $\pm 45°$ away from the studied view. When two views with an angular disparity of 75° were studied, recognition was constantly above 95% within the whole range spanned by the two views. This study is important for view-based theories for three reasons. First, the view-tuned cells found may be the neural realization of view-based functions of recognition performance. Second, the broad tuning property of the view-tuned cells suggested that the visual system needs not store a huge number of views of an object for view-invariant recognition. Third, the near-to-perfect generalization to views in between 75°-apart training views suggested the possibility that a small number of view-tuned cells act as a group and are sensitive to a novel view which activate each of the cells to a certain extent.

The suggestion that a population of cells is employed to represent an object at a particular view has been provided by Fujita, Tanaka, Ito, and Cheng (1992) and Wang, Tanaka, and Tanifuji (1994). Fujita et al. (1992) discovered the columnar organization in anterior IT by recording cell responses through penetrations perpendicular and oblique to the cortical surface. They found that cells within a column were responsive to more similar complex features compared with cells in different columns. Wang et al. (1994) further found that adjacent columns in IT representing similar features, such as different views of a face, overlapped with each other. In one of the experiments, they first identified through single-cell recording five cells that were most responsive to front faces or profiles. Then they adopted an optical imaging procedure to identify the areas with increased neuronal activities when the same face stimuli were introduced again. They discovered that, when the face stimuli changed from the left profile through 45°- and front faces to the right profiles, the center of the active areas also systematically shifted in one direction. Also, the active areas, each corresponding to a column about 300 to 400 μm wide, overlapped with each other and only occupied a total area of 800 μm. This suggested that different views of an object activated different but overlapping groups of cells in IT.

View-Invariant Recognition through Pooled Activation

With the above-mentioned findings, it is reasonable to postulate that when an object is studied at one view, it is stored as a view-specific representation involving a cluster of cells in IT. View-invariant recognition is likely the result of cooperation of a number of view-specific representations. A popular description of the mechanism of such view combination, hereafter referred to as the pooled activation account, has emerged and was developed in the past ten years (Bricolo, Poggio, & Logothetis, 1997; Perrett, Oram, & Ashbridge, 1998; Poggio & Edelman, 1990). Poggio and Edelman (1990) described a generalized radial basis function (GRBF) network, which uses information specific to training views of an object to generalize recognition to novel views. In a special case of such a network, the radial basis function (RBF) network, hidden units are present and the number of units equals the number of training views presented to the network. The activation of each of these view-tuned units, or non-linear receptive fields, is determined by the difference between an object view and its preferred view. When a novel view is introduced, it activates different view-tuned hidden units to different extents. The activities of different units are then pooled together by weighted summation and fed to a view-invariant, object-specific output unit. This model predicts that the response of an object recognition system to an object view is determined by the similarity between the current view and the views experienced before. It also predicts that a novel view has to lie within the range spanned by the studied views in order to activate sufficient pooled responses for its recognition. Bricolo et al. (1997) further developed this model by specifying the inputs to and organizations of the view-tuned hidden units, and obtained simulation results consistent with some of the behavioral data of Logothetis and Pauls (1995). Riesenhuber and Poggio (1999) integrated this type of view-tuned unit organization into a larger hypothesized network of visual object recognition invariant to size, location, and view changes.

Perrett et al. (1998) also proposed similar mechanisms of achieving view invariance as Poggio and Edelman (1990), but in neurophysiological terms. Their theory is based on Logothetis and Paul's (1995) results and their own findings of primate neural responses towards different views of head and body parts (Oram & Perrett, 1992; Perrett et al., 1991). They hypothesized that populations of cells are broadly tuned to particular object views. A novel view causes different levels of activity (i.e., firing rate) in different cell populations, depending on the similarity between the novel view and the preferred view of each population. As long as different views of an object have been associated to the same response, cells in the higher level sum up activity of individual cell populations over time, and the time for this summed activity to reach a threshold level determines the recognition time. This process is very similar to that described in Edelman and Poggio (1990) in which individual receptive field activity is combined linearly. Views between the studied views, according to Perrett et al.'s (1998) theory, should cause more activation of individual neuronal populations, leading to shorter time for the convergent output to reach the threshold level, and thus faster reaction time. Although no explicit prediction about recognition accuracy has been discussed, it is plausible that greater convergent activity is more distinguishable from the noise or background neuronal activity and should contribute to fewer errors.

Other theories of view combination include linear combination model and mental rotation theory. Ullman's (1996, 1998) linear combination model suggested that mathematically it is possible to generate all the views with experience of only a few views of a rigid object without self-occlusion across depth rotation. The general idea is that a 3-D object can be represented by a limited number of 2-D views of the object. When a novel object view is received, linear combination of the stored views for that object would occur to generate a new view for matching with the novel view. This model predicts that with two views of an object studied, linear combination should be equally capable of producing a new view resembling an internal or external view, and thus both internal and external facilitation should occur. Another theory is a mental rotation account which suggests that a novel view of a familiar object is recognized by rotating the nearest experienced view to match with the novel view (Tarr, 1995). According to this theory, the time required to recognize a novel view would be longer when the angular difference between the novel view and the nearest studied view gets larger. Increasing the number of studied views would not improve recognition of a novel view, if the angular difference is not changed. These two theories have received little behavioral support, and the present study helped to test the pooled activation account against them.

Empirical Studies about View Combination

Several human behavioral studies directly addressed the problem of how information from multiple views of an object can be combined for recognition of novel object views. Though inconclusive, they provide general support for the pooled activation account.

Bülthoff and Edelman (1992) investigated the pattern of generalization from familiar to unfamiliar views of amoeboid and wire-like objects (Figure 1a). Their experiment consisted of a series of alternating study and test trials. During each study trial, a visually novel object (amoebae-like or wire-like) was shown to the participants in 2-D images from two viewpoints which were 75° apart and oscillated for 15° along the vertical axis (Figure 1b). The oscillations produced a kinetic depth effect providing 3-D information of the object shape. During each test trial, a static image of either the just-viewed object or a distractor was presented in INTER views (within the inner 75° range spanned by the two training views), EXTRA views (within the outside 285° range spanned by the two training views), or ORTHO views (rotated around an orthogonal horizontal axis). Results showed the highest error rate for ORTHO views, followed by EXTRA views, which in turn was higher than the INTER views. With similar objects studied by monkeys, Logothetis and Pauls (1995) also found complete generalization to internal views with training views 75° apart and no extrapolation.

Insert Figure 1 here

The better generalization found for internal than external views was claimed to provide support for adoption of view combination mechanisms like

RBF morphing (Poggio & Edelman, 1990). It is interesting to note that similar results were found for the amoeboid objects, which were susceptible of self-occlusion across different viewpoints, and the wire-like objects, the vertices of which were visible at all views presented. However, there are two main problems in the experiments. First, there was not any baseline condition where only one view was studied. It is not known whether there is a benefit from studying two views rather than one. Second, the error rates of the experiments were so high that it is doubtful whether the task involved was similar to daily recognition (error rates as high as 70 % were observed in one of the ORTHO conditions).

Srinivas and Schwoebel (1998) probed whether presentation of two temporally separated views of an object leads to generalization to a third view outside the range spanned by the two studied views (external views). In general, the experiments involved a study phase in which a series of two consecutive views of each object were presented. Sometimes the 80° and 110° views of an object were presented consecutively (80°-110° condition), while at other times either a 0° view (0°-0° condition) or an 80° view (80°-80° condition) was presented twice consecutively. Participants made symmetry decisions for a set of bilaterally symmetric or asymmetric objects. Then in the *test phase* some participants made symmetry judgments for studied or novel objects while others performed an old/new recognition memory task. Only the 0° view was introduced for all objects. Results showed faster reaction time for both the 0°-0° and 80°-110° conditions than the 80°-80° condition with symmetric objects, but not asymmetric ones. In other words, studying both 80° and 110° views resulted in better recognition at the 0° view than studying the 80° view only. This generalization to a third external view occurred only when the object was symmetric.

Although the above study revealed external facilitation for symmetric objects (i.e., better generalization to an external view with two views studied), the results seem to be compatible with the pooled activation account. In Logothetis and Paul's (1995) study, monkey recognition performance of some objects was equally good at the studied view and also at a view 180° away from it. This view resembled the mirror image of the studied view. In the monkey's IT, 5 out of the 71 neurons showed the maximum responses towards a certain view, as well as the view resembling its mirror image. Therefore, it is reasonable to postulate that experience of the 80° and 110° views of symmetric objects developed sensitivity not only to these two views, but also to the two other views resembling the mirror image of the two studied views. Facilitation thus occurred for views lying among the studied views and their mirror-image-like views, leading to better performance at the 0° view. For asymmetric objects, sensitivity developed only to the 80° and 110° views and little generalization outside that range (e.g., 0°) was present.

Using symmetric objects only, Schwoebel and Srinivas (2000) further studied how one's ability to combine two object views is affected by similarity and temporal separation between views. Twelve views 21 bilaterally symmetric objects were generated by rotation around the vertical axis such that neighboring views were 30° apart. Perceived view similarity was then measured by asking participants to rate on a seven-point scale the similarity between different views, and then by analyzing the ratings using multidimensional scaling (MDS)[1]. After that, similar tasks as those in Srinivas & Schwoebel (1998) were introduced.

During the study phase, participants needed either to invent a function for each of the study objects or to judge whether the objects were more like a tool than a support. One or two views were presented for each object. Then in the test phase participants were asked to answer with key pressing whether each object presented had been studied before or not. One important result was that better generalization to novel views (e.g., a 90° view) occurred when the two studied views were dissimilar (e.g., 30° and 210°) than when they were similar (e.g., 0° and 30°). This suggests that any view combination account have to address the effect of similarity between studied views on later generalization to novel views.

Kourtzi and Shiffrar (1999) conducted four experiments to search for generalization to internal and external views. In a typical experiment, participants saw consecutively two prime views of an object at the beginning of each trial. Then two subsequent targets were shown simultaneously and participants had to decide whether the two targets matched one another. The targets sometimes depicted the same object as the prime, but in a different view (internal or external). Sometimes the target objects were different from the prime ones. Both objects with and without self-occlusion across depth rotation were tested in the experiment. Results showed generalization to internal views only when the prime views were 60° but not 120° apart, while generalization to external views was found for prime views separated by 60° in two out of four experiments. Although it is difficult to explain the inconsistency of results, the presence of internal and external facilitation for object rotations with and without self-occlusion is worth consideration.

<u>The Present Study</u>

The present study is a starting point of an investigation of how our visual system generalizes from familiar to novel views of objects. A sequential matching task was used in all four experiments (Figure 2). During each trial the participant was required to study one view or two views of an object, and then judge whether a test image depicted the same object as that just shown.

Insert Figure 2 here

A specific prediction of the pooled activation account was tested. According to this account, generalization to internal views should be better when the objects involved are not subject to self-occlusion across depth rotation than when self-occlusion occurs. The difference between occluded and non-occluded rotations lies mainly in the width of the tuning ranges of individual view-tuned neuronal populations, using the terms of Perrett et al. (1998). In behavioral studies, a greater recognition cost is generally found when a novel object view has different visible features from the familiar view of the same object (Biederman & Gerhardstein, 1993; Hayward & Tarr, 1997). For occluded rotation, individual view-tuned neuronal populations are sensitive to a smaller range of views than those for non-occluded rotation. If the two studied views of an object with occluded rotation are too far away from each other, their corresponding view-tuned neuronal populations are less likely to be sensitive to the whole range of views between the studied views. Any internal view will then activate mainly one of the neuronal populations only, leading to little internal facilitation. For non-occluded rotation, on the contrary, there is a larger chance that the neuronal populations of two far-apart studied views remain considerably sensitive to all internal views. An internal view will be more likely to activate both populations and internal facilitation will be more evident. Therefore, the angular disparity between the two studied views was varied across experiments. For amoeboid objects the angular difference between studied views was set at 40°, or 60°. For geon objects, the difference could be 56°, or 84°. It was predicted that, for non-occluded rotation, internal facilitation would occur even when the studied views were far apart; for occluded rotation, internal facilitation would be less obvious as the studied views became further apart.

Amoeboid objects similar to those in Bülthoff and Edelman's (1992) study and geon objects similar to those in other behavioral studies were used to increase generalizability of results. In Srinivas and Schwoebel's (1998) study, external facilitation was found only for symmetric objects but not asymmetric ones. As mentioned before, such external facilitation may actually be internal facilitation between the studied view and a virtual view of the mirror image of a studied view. To avoid bilateral symmetry (which complicates the comparison between internal and external facilitation), all objects were asymmetric. Some objects had features occluded across different views, while others had all protrusions visible at all presented views.

Experiment 1: Amoeboid Objects with Studied Views Separated by 40°

Experiment 1 was based on Bülthoff and Edelman's (1992) study, with several important modifications. One was a direct comparison between depth-rotation that did and did not result in self-occlusion. Second, recognition performance following study of two views was compared to a condition in which only one object view was studied. Third, task difficulty was adjusted in a pilot study to reduce the error rates. This not only made the task more similar to efficient daily object recognition, but also enabled both error rate and reaction time data to be analyzed.

It was expected that, for both non-occluded and occluded rotations, when only one view was studied, subsequent recognition of a novel view would deteriorate with the increase in angular disparity from the studied view. When two views were studied, it was predicted that this disadvantage for novel views would disappear at views within the range spanned by the studied views (internal facilitation), but not outside the range.

Method

Participants

Thirty students at the Chinese University of Hong Kong participated in the experiment with a payment of HK$60. All reported normal or corrected-to-normal eyesight.

Apparatus & Stimuli

A G3 and two iMac computers with a high-resolution monitor were used. Presentation of stimuli was controlled by RSVP software (Williams & Tarr, No date). The objects used in the experiment were created and rendered by Carrara software.

All objects were composed of a central sphere with six irregular protrusions of different lengths and pointing directions (Figure 3a). There were 20 objects in the *non-occluded rotation set* with all protrusions visible across the views shown, and 20 objects in the *occluded rotation set* with some protrusions visible at some presented views only. The objects were rendered with realistic lighting and shading on a black background, and given a yellow plastic texture. They were about 9 cm large, spanning a visual angle of about 6.9° for a participant sitting about 75 cm away from the monitor. Static images were created for five views of each object to be used in the test phase. The five views (views 1 to 5) were created by rotating the rendering camera around the vertical axis passing through the center of the central sphere of each object. An angular difference of 20° existed between adjacent views.

Insert Figure 3 here

Two movie clips, each lasting for 4.5 seconds, were created for each object for presentation in the study phase. Each clip either depicted the target object at one view (view 2 or view 4), or at two views consecutively (views 2 and view 4). The objects shown in the movie engaged in 5° oscillations around the vertical axis to provide participants with information about the 3-D structure of the objects.

Design

There were three within-subject variables (rotation type × study condition × test view). The occluded and non-occluded sets formed the two levels of rotation type. The study condition variable consisted of the single condition, where only view 2 or view 4 was presented in the study phase[1], and the double condition, where both view 2 and view 4 were shown. For test view, view 2 and/or view 4 were regarded as studied views[2], while view 3 as internal views, and view 1 and/or view 5 as external views. Table 1 shows the experimental design.

Insert Table 1 here

In the single study condition, higher error rate and reaction time was

[1] For the single study condition, view 2 was always presented for half of the participants to study whereas view 4 was always presented for the other half.

[2] For the single study condition, only the presented view (view 2 or view 4) was counted as the studied view. When view 2 was the studied view, view 1 and view 3 were the external and internal views respectively. For the double study condition, both view 2 and view 4 were presented, so they were regarded together as the studied views. Views 1 and 5 were regarded as the external views, while view 3 as the internal view.

predicted at the internal and external views than at the studied view. In the double study condition, internal but not external facilitation was predicted. Internal facilitation would be indicated by the disappearance of the studied view's advantage over the internal view, i.e., similar error rate and reaction time at the two views. Lack of generalization to external views would be indicated by the higher error rate and reaction time at the external views than the studied views.

Procedure

The five views for each of the 40 objects (20 in the occluded set and 20 in the non-occluded set) formed a total of 200 trials with target views. An equal number of distractor trials were created by using each target trial and replacing the target view with the corresponding view of another object in the same set. Inclusion of the same number of distractors led to a total of 400 trials.

Each trial consisted of a study and a test phase. In the 4.5-second study phase, sequential presentation of two views (in the single study condition, the two views were the same; in the double study condition, they were different) of a target object was shown for participants to memorize. A mask was then presented for 500 ms, followed by the test phase with a static view of either the target or a distractor. Participants had to press the appropriate key ("1" for target and "2" for distractor) accurately and as fast as possible. Upon the participant's response, the test view disappeared and the second test trial began. The trials were randomly presented and separated in five blocks of 80 trials each. Four practice trials were provided before the experimental trials.

Results

Participants responded correctly in 84% of the trials on average. Because of a program error the data from two participants were discarded. For both error rate and RT analyses only the target trials were included. From RT analysis 14.2% of the target trials were further discarded because of response error, and on top of that 0.9% of the trials were discarded because the RT was shorter than 250 ms or longer than 4000 ms in these trials. The resulting mean RT was 1017 ms (SD = 577 ms).

Four 2 × 3 analyses of variances (ANOVAs) were conducted, for non-occluded and occluded rotations, with the study condition (single, double) and test view (external, studied, internal) as independent variables, and error rate and RT as the dependent variables. Results are shown in Figure 4.

Insert Figure 4 here

Non-occluded rotation

Clear internal facilitation but not external facilitation was found for the non-occluded rotation set of objects (Figure 4a). It can be seen that when one view was studied, performance was worse at the internal and external views than the studied views. When two views were studied, performance at the external views remained worse than at the studied views, but there was little difference between performance at the internal and studied views.

Statistical analyses confirmed the above descriptions. With error rate as the dependent variable, the main effect of test view was significant [$F(2,54) = 9.931$, $p < .001$], whereas that of study condition was not [$F < 1$]. There was also a significant interaction between test view and study condition [$F(2,54) = 3.467$, $p = .038$]. In the single study condition, errors were lower at the studied view than at the external [$t(27) = 2.375$, $p = .025$] and internal views [$t(27) = 1.910$, $p = .067$, marginally significant]. In the double study condition, however, errors at the internal view were no longer different from those at the studied views [$p > .20$], while the error rate at the external views remained higher than at the studied views [$t(27) = 4.449$, $p < .001$].

Similarly, with RT as a dependent variable, the main effect of test view was significant [$F(2,54) = 14.949$, $p < .001$]. The main effect of study condition and the interaction between test view and study condition were not significant [Fs <1]. RT at the external view was consistently higher than at the studied view in both single [$t(27) = 2.885$, $p < .008$] and double study conditions [$t(27) = 5.529$, $p < .001$]. There was no difference between RT at the studied and internal views in both single and double study conditions [ts < 1].

<u>Occluded rotation</u>

Results for the occluded rotation set of objects were similar to the non-occluded rotation set (Figure 4b). Again worse performance occurred at the external and internal views than at the studied view when only one view had been studied. When two views had been studied, performance at the internal view became even better than at the studied views, an indication of internal facilitation. Recognition at the external view was still less accurate than at the studied view, though reaction time was similar at the two views.

With error rate as the dependent variable, there were significant main effects of study condition [$F(1,27) = 7.549$, $p = .011$] and test view [$F(2,54) = 17.619$, $p < .001$], and a significant interaction [$F(2,54) = 9.419$, $p < .001$]. For the single condition, responses at the studied views were more accurate than at the external [$t(27) = 3.667$, $p = .001$] and internal views [$t(27) = 4.473$, $p < .001$]. However, for the double study condition, responses at the internal view were more accurate and faster than that at the studied views [$t(27) = 2.273$, $p = .031$]. Responses at the external views were still less accurate than that at the studied views [$t(27) = 3.207$, $p = .003$]. One more point to note is that responses at the studied view were less accurate in the double than in the single study conditions [$t(27) = 5.157$, $p < .001$]. This showed that it was easier to memorize one single view than to memorize two views that did not share exactly the same visible protrusions.

With RT as the dependent variable, there were significant main effects of study condition [$F(1,27) = 9.020$, $p = .006$] and test view [$F(2,54) = 3.007$, $p = .058$, marginally significant], and a significant interaction [$F(2,54) = 6.175$, $p = .004$].

In the single study condition, responses to the studied view were faster than at the external [$t(27) = 3.227, p < .003$] and internal views [$t(27) = 3.298, p < .003$]. In the double study condition, however, responses at the internal view were faster than that at the studied views [$t(27) = 2.018, p < .054$, marginally significant], while responses to the external and studied views were similar [$t < 1$]. Again, there was a higher error rate for the studied views in the double study condition than one in the single study condition [$t(27) = 5.804, p < .001$].

Discussion

Results of Experiment 1 showed internal but not external facilitation for both non-occluded and occluded rotations. This is consistent with and compliments the findings of Bülthoff and Edelman (1992). The inclusion of the single study condition showed that generalization to external views did not improve with two views studied compared with the case when only one view was studied. Also, the higher accuracy rate in Experiment 1 gives greater credibility to the results.

According to the pooled activation account, with two views studied, the internal view was close to both studied views (view 3 was close to both view 2 and view 4) while the external view was close to one of the studied views only (view 1 was close to view 2 only, and view 5 was close to view 4 only). Therefore the internal view caused higher activation in the cells corresponding to the two studied views, leading to higher accumulated responses and thus better performance. This activation appeared not to be modulated by occlusion.

Experiment 2: Amoeboid Objects with Studied Views Separated by 20° and 60°

In Experiment 2 two issues were further examined. The first concerned whether there is any change in internal facilitation found for non-occluded and occluded rotations with increase in the angular disparity between the studied views. When the two studied views do not share all the important features, i.e., there is self-occlusion across depth rotation, their corresponding view-tuned neuronal populations are likely to have narrower tuning ranges. So when the two studied views were far apart, the two neuronal populations are not likely to be sensitive to all the internal views. An internal view thus tends to activate only one of the view-tuned neuronal population, leading to very little internal facilitation. Wider tuning ranges, however, exist when no self-occlusion occurs across depth rotation. Thus, in this case an internal view is likely to activate both populations to a large extent, causing more internal facilitation. Therefore, less internal facilitation was predicted with far-apart studied views for objects with self-occlusion across depth rotation than those without. To test this prediction, in Experiment 2, angular disparities were enlarged from 40° in Experiment 1 to 60°. It was expected that internal facilitation across occluded rotations would be less obvious with increased angular disparity between the two studied views.

The second issue this experiment explored was whether generalization will still improve after studying two very similar views. There were three study conditions in Experiment 2. In the single study condition, only one object view (e.g., view 2) was studied. In the double-adjacent study condition, there were two close studied views (e.g., views 1 and 2), while in the double-separated study condition, the two studied views were farther apart from each other (e.g., views 2 and 5). Of particular interest was the recognition performance for the novel view 3 in all the three study conditions. It should be noted that in the double-adjacent study condition, view 3 was an external view 20° away from the nearest studied view (e.g., view 2) and 60° away from the further studied view (e.g., view 1). In the double-separated condition, view 3 was an internal view 20° away from the nearest studied view (e.g., view 2) and 60° away from the further studied (e.g., view 5). If the similarity between the studied views is not an important factor of view combination, then generalization to the novel view 3 should be comparable in the double-adjacent and double-separate study conditions, because the angular disparities between the novel view 3 and the two studied views were the same. If, instead, it is important to avoid studying two nearby views for better improvement in generalization then the resulting generalization to the novel view 3 in the double-adjacent condition would be similar to the case when only one view was studied, i.e., no external facilitation would be found.

Method

Participants

Fourty-seven university students participated in the experiment for partial fulfillment of a course requirement. All reported normal or corrected-to-normal eyesight, and had not participated in any of the other three experiments. Twentyfour participants took part in a half-hour session and were presented with the non-occluded object set only, while 22 also took part in a half-hour session and were presented with the occluded object set only. One participant took part in both sessions and was presented with both sets of objects.

Apparatus & Stimuli

Apparatus and stimuli were the same as in Experiment 1 except for two differences. First, 60 amoeboid objects were used, with 30 objects in the occluded rotation set and 30 in the non-occluded rotation set. Second, static instead of oscillating stimuli were used in the study phase.

Design

Rotation type (non-occluded, occluded), study condition (single, double-adjacent, double-separated), and test view (nearest studied, novel) formed the three variables in the present study. As shown in Table 1, for the single study condition, only view 2 or view 4 was presented in the study phase; for double-adjacent study condition, views 1 and 2 were studied in some trials while views 4 and 5 were studied in other trials; finally, for the double-separated study condition, both views 2 and 5 were studied in some trials while both views 1 and 4 were studied in other trials. Of particular interest was the difference in performance at the novel view 3 (hereafter named as the novel view) and its closest studied view (hereafter named as the nearest studied view). In the single study condition, this difference corresponded to the viewpoint cost of recognizing view 3 with view 2 or view 4 studied. In the double-separated condition, the absence of the difference would indicate an internal facilitation. In the double-adjacent condition, the absence of the difference would indicate an external facilitation. According to the pooled activation account, it was predicted that internal facilitation would be more evident for the non-occluded object set than for the occluded object set. In addition, if the two studied views in the double-adjacent condition were too close to each other to form two representations, then there should not be any external facilitation.

Procedure

In each trial, during the study phase, a fixation cross appeared for 1000 ms. The first study view was then shown for 300 ms, followed by a 750-ms blank and the second study view for another 300 ms. A blank was presented for 500 ms, followed by a 510-ms mask and then the test phase with a static view of either the target or a distractor. Participants were asked to press the appropriate key ("1" for target and "2" for distractor) accurately and as fast as possible. Upon the participant's response, the test view disappeared and the second test trial began. The trials were randomly presented and separated in six blocks of 100 trials each. Six practice trials were provided before the experimental trials.

Results

Participants were correct in 79 % of the trials. For both error rate and RT analyses only the target trials were included. From RT analysis 18.9% of the target trials were further discarded because of response error, and on top of that 0.98% of the trials were discarded because the RT was shorter than 250 ms or longer than 4000 ms in these trials. The resulting average reaction time of 1151 ms (SD = 583 ms).

For both the non-occluded and the occluded object sets, two 3 × 2 ANOVAs were conducted with study condition (single, double-adjacent, double-separated) and test view (nearest studied, novel) as independent variables, and error rate and RT as dependent variables.

Insert Figure 5 here

Non-occluded rotation

Internal facilitation was present whereas external facilitation was not for objects in the non-occluded rotation set (Figure 5a). It can be seen that performance at the novel view was worse than at the nearest studied view in the single study condition, as well as when two close views were studied, which indicates a lack of external facilitation. However, when two far-apart views were studied, subsequent recognition efficiency at the novel view and the nearest studied view was about the same, which is a sign of internal facilitation. The increase in the angular disparity between the studied views exerted little effect on generalization to the internal views.

Statistical tests confirmed the above descriptions. With error rate as the dependent variable, the main effect of study condition was significant [$F(2,48) = 5.472, p < .007$]. So was the main effect of test view [$F(1,24) = 11.888, p < .002$]. The interaction between them was not significant [$F(2,48) = 1.412, p > .25$]. Subsequent *t*-tests showed that, for the single study condition, more accurate responses were found at the nearest studied view than at the novel view [$t(24) = 3.335, p = .003$]. Similarly, for the double-adjacent study condition, more accurate responses were found at the nearest studied view than the novel view [$t(24) = 2.735, p = .012$]. For the double-separated study condition, however, no differences were found between the nearest studied and the novel views [$t < 1$], which is a sign of internal facilitation.

With RT as the dependent variable, the main effect of study condition was not significant [$F < 1$]. There was a significant main effect of test view [$F(1,24) = 6.164, p < .02$] and an interaction between study condition and test view [$F(2,48) = 6.385, p < .003$]. For the single study condition, faster responses were found at the nearest studied view than at the novel view [$t(24) = 5.132, p < .001$]. For the double-adjacent study condition, differences were still found between the nearest studied view and the novel view [$t(24) = 2.255, p = .034$]. This again showed that studying two close views contributed little to studying a novel view outside. In the double-separated study condition, however, response to the nearest studied view was not different from responses to the novel view [$t(24) = 1.017, p > .31$].

<u>Occluded rotation</u>

Unlike the results of Experiment 1 and the non-occluded condition of this experiment, internal facilitation was not found for objects in the occluded rotation set (Figure 5b). Performance at the nearest studied view was better than at the novel view when only one view was studied. This studied view advantage still remained when two views were studied, no matter whether they were close to or far apart from each other.

With error rate as the dependent variable, a significant main effect was found for test view [$F(1,22) = 28.793$, $p < .001$]. There was no significant main effect of study condition [$F(2,44) = 2.292$, $p > .11$], nor was there an interaction between study condition and test view [$F(2,44) = 1.022$, $p > .36$]. Subsequent *t*-tests showed that responses at the novel view were less accurate than at the nearest studied view in all three study conditions [**single**: $t(22) = 3.590$, $p < .002$; **double-adjacent:** $t(22) = 4.511$, $p < .001$; **double-separated**: $t(22) = 3.239$, $p < .004$]. Therefore, whether studying one view, two views which embedded view 3, or two views which did not embed view 3 made no difference in later generalization of recognition to view 3.

RT analyses showed similar results. The main effect of test view was significant [$F(1,22) = 6.385$, $p < .02$]. There was no main effect of study condition [$F(2,44) = 2.128$, $p > .13$] or a significant interaction [$F < 1$]. Strangely, RT was higher at the novel view than at the nearest studied view in the double-separated study condition only [$t(22) = -2.829$, $p < .01$] but not in the other conditions [*ps* $> .14$]. This may be just caused by differences in RT variability in the three study conditions, with regard to the similar values of the RT difference between the nearest studied and novel views in the three conditions.

Discussion

Two main findings are worth our attention. First, internal facilitation was shown in the double-separated condition for the non-occluded object set, which is consistent to the results of Experiments 1. For the occluded object set, however, internal facilitation was not found, which is different from the results of Experiment 1. This may have been because the angular disparity between the studied views (60°) was larger than that in Experiment 1 (40°). Such an explanation supports the suggestion that the view-tuned neuronal populations of the studied views have wider tuning ranges when objects have no self-occlusion across depth rotation. Indeed, in both experiments, when only one view was studied, recognition at test views 60° away from the studied view was still quite accurate for the non-occluded object set (**Experiment 1:** 72%; **Experiment 2:** 64%). Recognition accuracy for the occluded object set, however, dropped considerably with a test view separated from the closest studied view by 40° (**Experiment 1:** 65%; **Experiment 2:** 59%) and reached chance level with a test view rotated by 60° (**Experiment 1:** 49%; **Experiment 2:** 50%). Therefore, considerable sensitivity towards internal views was still retained for the non-occluded object set even when the angular disparity was 60°. For the occluded object set, however, not much sensitivity was left with a 60° disparity. It is plausible that an internal view more likely activates both view-tuned neuronal populations for the non-occluded object set, resulting in internal facilitation.

Second, no external facilitation was found when two close views were studied in the double-adjacent condition, even though the angular disparity between the novel view and the two studied views was as favorable as in the double-separated study condition in which internal facilitation was shown. The resultant generalization to novel view 3 in the double-adjacent condition is similar that in the single-study condition. This is consistent with the findings of Schwoebel and Srinivas (2000), in which improved generalization was not found when the two studied views were perceived to be very similar to each other. With regard to this finding and the current results, it is doubtful whether, after cells have developed sensitivity to the first studied view, another group of cells are also modified to develop sensitivity to the second studied view.

Experiment 3: Geon objects with Studied Views Separated by 56°

As in Experiment 1, the aim of Experiment 3 was to probe the advantage of studying two views over one in generalizing recognition to novel views. There were three main differences between this experiment and Experiments 1 and 2. First, geon objects composed of geon-like volumes (Biederman, 1987; Biederman & Gerhardstein, 1993) were used instead of amoeboids (see Figure 5b). For the latter, protrusions were all similar in shape and different only in terms of length and thickness, and thus the main difference among the amoeboid objects thus lay in the metric properties and spatial arrangement of protrusions. The geon objects had a central part informative of the object view (i.e., the central part appeared differently at different views), and geon protrusions with distinct shapes.

According to the distinctions made by Biederman and Bar (1999), individual amoeboid objects would be regarded as different in metric properties affected by depth rotations. Differences among individual geon objects, however, would depend on the non-accidental properties (e.g., a vertex), which are generally unaffected by changes in viewpoint. It would be interesting to see if similar viewpoint dependence and view combination patterns are observed in the geon objects.

The second difference of Experiment 3 was that the viewpoint difference between adjacent views was increased from 20° to 28°. Third, more features were occluded during depth rotation for the objects in the occluded rotation set. These were designed to make the viewpoint dependent response in the single study condition and any internal and external facilitation in the double study condition more obvious. Again, internal but not external facilitation was expected when two views were studied, according to the pooled activation account.

Method

Participants

Twenty-three university students at the Chinese University of Hong Kong participated in the experiment either for fulfillment of a course requirement or for a payment of HK$ 50. Again all reported normal or corrected-to-normal vision, and none had participated in any of the other three experiments. They all viewed both occluded and non-occluded sets of objects.

Apparatus & Stimuli

The same computer, monitor, and object creation and presentation softwares were used as in Experiments 1 and 2.

Each of the objects had a central volume with six geon protrusions of different shapes and pointing directions. There were 20 objects in the *non-occluded rotation set* with all protrusions visible across the views shown, and 20 objects in the *occluded rotation set* with some parts visible at some presented views only. The objects were again rendered with realistic lighting and shading on a white background, and given a red wooden texture. They were about 4 cm large, extending a visual angle of 3.1° for a participant sitting 75 cm away from the monitor. Static images were created for five views of each object to be used in test trials (views 1 to 5, with an angular difference of 28° between adjacent views).

Design

The design was the same as that in Experiment 1, with three within-subject variables (rotation type × study condition × test view). The occluded and non-occluded sets formed the two levels of rotation type. The study condition variable consisted of the single condition, where only view 2 or view 4 was presented in the study phase, and the double condition, where both view 2 and view 4 were shown. For the test view, view 2 and/or view 4 were regarded as the studied views, while view 3 was the internal views, and view 1 and/or view 5 were external views.

Procedure

Similar procedures as those in Experiment 1 were adopted, with a major change that during the study phase, static instead of oscillating views were presented. In each trial, during the study phase, a fixation cross appeared for 1000 ms. The first study view was then shown for 300 ms, followed by a 750-ms blank and the second studied view for another 300 ms. A mask was then presented for 510 ms, followed by the test phase with a static view of either the target or a distractor. Participants had to press the appropriate key ("1" for target and "2" for distractor) accurately and as fast as possible. Upon the participant's response, the test view disappeared and the second test trial began. The trials were randomly presented and separated in five blocks of 80 trials each. Six practice trials were provided before the experimental trials.

Results

On average participants were correct in 81 % of the trials. The data from three participants were discarded because two of them had near-chance-level performance throughout the experiment (ERs = .45 and .41), while one had near-chance-level performance in one of the blocks (ER = .44 in one of the blocks). For both error rate and RT analyses only the target trials were included. From RT analysis 13.1% of the target trials were further discarded because of response error, and on top of that 0.55% of the trials were discarded because the RT was shorter than 250 ms or longer than 4000 ms in these trials. The resulting mean RT in all the trials was 1103 ms (SD = 548 ms).

For both non-occluded and occluded sets of objects, two 2 × 3 ANOVAs were conducted with study condition (single, double) and test view (external, studied, internal) as independent variables and with error rate and RT as dependent variables

----

Insert Figure 6 here

Non-occluded rotation

Internal facilitation but not external facilitation was found for the non-occluded object set (Figure 6a). It can be seen that when one view was studied, performance was worse at the internal and external views than at the studied views. When two views were studied, performance at the external views remained worse than at the studied views, but there was little difference between performance at the internal and studied views. This loss of viewpoint cost for the internal view with two studied views indicates an internal facilitation.

Statistical tests confirmed the above descriptions. With error rate as the dependent variable, the main effect of test view was significant [$F(2,38) = 7.462$, $p < .002$]. The main effect of study condition was not significant [$F(1,19) = 2.307$, $p > .14$]. The interaction effect approached significance [$F(2,38) = 2.981$, $p = .063$].

In both single and double study conditions performance at the external view was less accurate than at the studied view [**single:** $t(19) = 2.483$, $p < .03$; **double:** $t(19) = 3.111, p < .006$]. However, error rate at the internal view was higher than at the studied views only for the single study condition [$t(19) = 3.627, p < .002$] but not for the double study condition [$t(19) = 1.238, p > .23$]. This means that the studied view's advantage over the internal view became negligible when two views instead of one view were studied.

Similar results were obtained with RT as the dependent variable. The main effect of test view was significant [$F(2,38) = 18.269, p < .001$], whereas the main effect of study condition was not [$F < 1$]. A significant interaction effect between test view and study condition was found [$F(2,38) = 8.832, p < .001$]. Multiple *t*-tests showed that RT at the external view was higher than at the studied views for both single and double study conditions [**single:** $t(19) = 5.733, p < .001$; **double:** $t(19) = 4.447, p < .001$], which indicated no external facilitation. RT at the internal view was higher than at the studied views for the single condition [$t(19) = 4.432, p < .001$], but not for the double condition [$t < 1$], which is a sign of internal facilitation.

Occluded rotation

Little internal facilitation was found in the occluded object set (Figure 6b). Recognition at the external and internal views was worse than at the studied view when only one view was studied. When two views were studied, however, performance at the internal views became almost equally fast as that at the studied views. Since more error was still involved in the internal view than at the studied views, at most the results can only be treated as a weak sign of internal facilitation. Recognition at the external views remained worse than at the studied views, i.e., there was no external facilitation.

With error rate as the dependent variable, only the main effect of test view was significant [$F(2,38) = 14.970$, $p < .001$]. Neither the main effect of study condition [$F < 1$] nor the interaction effect was significant [$F(2,38) = 1.708$, $p > .19$]. Similar response patterns were found for both single and double study conditions, in that error rate was higher at the external than the studied views [**single:** $t(19) = 3.882$, $p < .001$; **double:** $t(19) = 3.390$, $p < .003$]. Also, error rate was higher at the internal than the studied views [**single:** $t(19) = 3.684$, $p < .002$; **double:** $t(19) = 3.101$, $p < .006$].

Some traces of internal facilitation were found in RT analyses. The main effect of test view was significant [$F(2,38) = 9.837$, $p < .001$]. The main effect of study condition was not significant [$F(1,19) = 1.585$, $p > .22$]. However, there was a significant interaction between test view and study condition [$F(2,38) = 5.452$, $p < .008$]. Responses at the external view were slower than at the studied view for both single [$t(19) = 2.264$, $p < .04$] and double study conditions [$t(19) = 4.789$, $p < .001$]. In contrast, the studied view's advantage over the internal view existed only in the single study condition [$t(19) = 5.025$, $p < .001$] but was only marginally significant in the double condition [$t(19) = 1.738$, $p > .09$].

Discussion

Results showed internal but not external facilitation for non-occluded and, less obviously, occluded rotations. This is consistent with the prediction of the pooled activation account, that the internal view is similar to both studied views, causing higher activation of individual view representations and higher accumulated responses and thus better performance. The more evident internal facilitation found for non-occluded rotation is consistent with the results of Experiment 2. When the two studied views share similar important features (non-occluded rotation), their view-tuned neuronal populations have greater sensitivity to all the internal views, compared to the condition when the studied views have different visible features (occluded rotation). An internal view is thus more likely to cause considerable activation in both neuronal populations in the non-occluded rotation case, causing more internal facilitation. One may ask why such difference between the occluded and non-occluded rotation was not found in Experiment 1. Reasons may include that the extent of self-occlusion and the angular disparity between adjacent views were greater than those in Experiment 1.

Experiment 4: Geon objects with Studied Views Separated by 28° and 84°

In Experiment 2 it was shown that when two views that were far apart had been studied, subsequent recognition at the novel view was better than when two close views had been studied. When the studied views were similar to each other, subsequent recognition performance was similar to the case in which only one view had been studied. In Experiment 4, the angular difference between neighboring views was increased so that the two studied views in the double-adjacent study condition looked less similar to each other. As in Experiment 2, there were three study conditions: the single condition where only view 2 or only view 4 was studied, the double-adjacent condition where views 1 and 2 were studied in some trials while views 4 and 5 were studied in other trials, and the double-separated condition where views 2 and 5, or views 1 and 4, were studied. It was expected that recognition performance would be better at the nearest studied view than at the novel view 3 with only one view studied. Of particular interest was whether generalization to the novel view 3 would improve with two similar studied views in the double-adjacent study conditions. If this was the case, then the difference in performance between the nearest studied view and the novel view 3 would also disappear, since view 3 would probably activate representations of the two close studied views to a similar extent as it would activate representations of two far-apart views in the double-separated study condition.

Method

Participants

Twenty-five university students at the Chinese University of Hong Kong participated in the experiment either for fulfillment of a course requirement or for a payment of HK$ 50. All reported normal or corrected-to-normal vision, and had not participated in the prior three experiments.

Apparatus & Stimuli

Apparatus and stimuli were the same as in Experiment 2 except that 60 geon objects were used, with 30 in the occluded rotation set and 30 in the non-occluded rotation set.

Design & Procedure

The design and procedure were the same as in Experiment 2.

Results

On average participants were correct in 79 % of the trials. The data from three participants were discarded, because the performance of two of them was near chance level (ERs = .50 and .40 respectively), and there was no hit trial in one of the conditions for the remaining participant. For both error rate and RT analyses only the target trials were included. From RT analysis 15.7% of the target trials were further discarded because of response error, and 0.71% of the trials were further discarded because the RT was shorter than 250 ms or longer than 4000 ms in these trials. The resulting mean RT in all the trials was 1146 ms (SD = 569ms).

For both non-occluded and occluded rotation sets of objects, two 2 × 3 ANOVAs were conducted with study condition (single, double-adjacent, double-separated) and test view (nearest studied, novel) as independent variables and with error rate and RT as dependent variables.

Insert Figure 7 here

Non-occluded rotation

Internal facilitation was present whereas external facilitation was not for objects in the non-occluded rotation set (Figure 7a). Performance at the novel view was worse than at the nearest studied view when only one view was studied, as well as when two close views were studied. This pattern indicates a lack of external facilitation. The similar response patterns in these two conditions suggest that only one view-specific representation was formed even though two views were experienced. However, when two far-apart views were studied, subsequent recognition efficiency at the nearest studied view and the novel view was about the same, which is a sign of internal facilitation. The increase in the angular disparity between the studied views exerted little effect on generalization to internal views.

Statistical tests confirmed the above descriptions. With error rate as the dependent variable, there was no significant main effect of study condition [$F < 1$]. There was a significant main effect of test view [$F(1,22) = 19.791$, $p < .001$], as well as an interaction between study condition and test view [$F(2,44) = 3.841$, $p < .03$]. For the single study condition, responses were more accurate at the nearest studied view than at the novel view [$t(22) = 4.084$, $p < .001$]. For the double-adjacent study condition, there were fewer errors at the nearest studied view than at the novel view [$t(22) = 3.950$, $p < .001$]. For the double-separated study condition, as expected, no error rate difference was found between the nearest studied view and the novel view [$t < 1$]. The error rate at the nearest studied view was higher for the double-separated condition than for the single condition [$t(22) = 2.533, p < .02$] and the double-adjacent condition [$t(22) = 2.772, p < .02$].

With RT as the dependent variable, there was a significant main effect of test view [$F(1,22) = 16.890, p < .001$]. There was no significant main effect of study condition [$F < 1$], and no interaction between study condition and test view [$F(2,44) = 1.550, p > .22$]. For the single study condition, responses were faster at the nearest studied view than at the novel view [$t(22) = 3.310, p < .003$]. For the double-adjacent study condition, the advantage of the nearest studied view over the novel view remained [$t(22) = 2.867, p > .009$]. For the double-separated study condition, however, the difference between the nearest studied view and the novel view only approached significance [$t(22) = 2.003, p = .058$]. Results again showed higher RT at the nearest studied view for the double-separated condition than the single condition [$t(22) = 2.960, p < .007$].

Occluded rotation

No evidence of interpolation was found for the occluded set of objects (Figure 7b). Recognition at the nearest studied view was consistently better than at the novel view no matter whether only one view had been studied, two close views had been studied, or two far-apart views had been studied.

With error rate as the dependent variable, significant main effects of study condition [$F(2,44) = 3.921, p > .03$] and test view [$F(1,22) = 44.913, p < .001$] were found. The interaction between study condition and test view was not significant [$F(2,44) = 2.120$, $p > .13$]. For the single study condition, responses were more accurate at the nearest studied view than at the novel view [$t(22) = 5.663$, $p < .001$]. The performance difference between the nearest studied view and the novel view remained for the double-adjacent study condition [$t(22) = 5.059$, $p < .001$] and the double-separated study condition [$t(22) = 2.954$, $p < .007$]. As with non-occluded rotation, more errors were involved in the nearest studied view in the double-separated condition compared with the two closer studied views in the double-adjacent condition [$t(22) = 2.712$, $p < .020$] and with the single studied view in the single condition [$t(22) = 3.793$, $p < .001$].

With RT as the dependent variable, there was a significant effect of test view [$F(1,22) = 38.046$, $p < .001$]. The main effect of test view approached significance [$F(2,44) = 2.692$, $p = .079$], while the interaction between study condition and test view was not significant [$F(2,44) = 1.367$, $p > .26$]. For the single study condition, responses were faster at the nearest studied view than at the novel view [$t(22) = 3.882$, $p < .001$]. Such difference between the nearest studied view and the novel view remained for the double-adjacent study condition [$t(22) = 6.330$, $p < .001$] and for the double-separated study condition [$t(22) = 2.608$, $p < .02$]. Again, it took longer to recognize the two studied views in the double-separated study condition than in the double-adjacent condition [$t(22) = 4.076$, $p < .001$] and in the single condition [$t(22) = 2.295$, $p < .04$].

Discussion

Internal facilitation for the non-occluded object set was observed in the double-separate condition, where the studied views were far apart (84°). However, for the occluded object set, little internal facilitation was found. These were consistent with results in the previous Experiments 2 and 3. With the same angular disparity it is harder to match the views of objects with self-occlusion since each view has unique features. Such a problem is worse with larger angular disparity. View generalization in the single study conditions in Experiments 3 and 4 was again examined. When only one view was studied, recognition at test views 84° away from the studied view was still quite accurate for the non-occluded object set (Experiment 3: 86%; Experiment 4: 82%). Recognition accuracy for the occluded object set, however, dropped much with test views rotated 56° (Experiment 3: 69%; Experiment 4: 69%) and 84° from the studied view (Experiment 3: 65%; Experiment 4: 66%)

One may question whether the lack of difference between the studied and novel views for non-occluded rotation in the double-separated condition was a result of the performance drop for the studied views. This interpretation will be inadequate if results for occluded rotation are also examined. For both occluded and non-occluded rotation, recognition of the studied view was generally worse in the double-separated than the other conditions. However, a difference between the novel and the nearest studied view was still found in the double-separated condition for occluded rotation. Therefore, the lack of difference for non-occluded rotation genuinely represented internal facilitation rather than artifact caused by the ceiling effect.

In the double-adjacent study condition of Experiment 4, although the angular disparity between the two studied views was enlarged to 28°, no external facilitation was obtained. Perhaps further increasing the angular disparity would make the two studied views different enough for any improvement in view generalization. However, the angular disparity between the external views and the two studied views would also be increased. Therefore, it is likely that view combination is mainly achieved for views within the range spanned by the studied views, not outside the range.

General Discussion

Interpolation by Pooled Activation

The findings of the present study support the pooled activation account of view combination in two ways. First, evidence for interpolation was found for asymmetric objects with or without occlusion across depth rotation. This is consistent with the prediction of the pooled activation account, which states that the view generalization advantage for studying two views instead of one lies in the ability of the visual system to combine responses of individual view-specific representations. As long as the two studied views were similar enough, the corresponding view-specific representations will have part of their generalization gradients overlap with each other. Any novel views in between the two studied views can thus activate both representations to a certain extent, and the combination of individual responses will result in faster and more accurate recognition. Since evidence was revealed only for interpolation but not extrapolation in this study with asymmetric objects, it is reasonable to believe that the external facilitation found with bilaterally symmetric objects in Srinivas and Schwoebel's (1998) study may be caused by interpolation between the experienced views and a generated virtual view resembling the mirror image of the experienced views.

The second support to the theory was provided by the finding that interpolation was more evident in objects without self-occlusion than objects with self-occlusion across depth rotation. Evidence for interpolation was found with the non-occluded object set no matter whether the angular disparity between the studied views was as small as 40° (with amoeboid objects in Experiment 1) or as large as 84° (with geon objects in Experiment 4). The evidence for interpolation in objects in the occluded set, however, became less conspicuous with increase in the angular disparity between the studied views. According to the pooled activation account, this was probably because the generalization gradients of the view-specific representations were narrower in objects with self-occlusion than objects without across rotation. With the same angular disparity between the studied views, novel views between the studied views of objects with occluded rotation were thus more likely to activate only one of the representations, or both representations to a lesser extent, leading to similar performance pattern as when only one view was studied. This explains why internal facilitation occurred for amoeboid objects with occluded rotation when the angular disparity between the studied views was 40° in Experiment 1, but not when it was increased to 60° in Experiment 2.

Another finding which supplements our understanding of the view combination process is the absence of extrapolation in Experiments 2 and 4. In these two experiments the angular disparity of the studied views was decreased such that the external view was close to both studied views. In this condition, the two studied views might be too similar to enable formation of two view-specific representations and thus produce any improvement in view generalization. Instead the preferred view of the existing representation might have been modified so that it was sensitive to both studied views. As Poggio and Edelman (1990) suggest, "When the number of basis functions is less than the number of views in the training set, the centers of the basis functions are also updated during learning… It is of interest that after training, the centers of the radial basis units correspond to views that are different from any of the training views" (p. 264). Schwoebel and Srinivas (2000) also presented a similar idea that the two studied views have to be different enough for good generalization to novel views.

It has been hypothesized that the visual system may somehow actively synthesize the appearance of all the views between the two studied views. Ullman's (1996, 1998) linear combination model suggested that mathematically it is possible to generate all the views with experience of only a few views of a rigid object without self-occlusion across depth rotation. The general idea is that a 3-D object can be represented by a limited number of 2-D views of the object. When a novel object view is received, linear combination of the stored views for that object would occur to generate a new view for matching with the novel view. This model predicts that with two views of an object studied, linear combination should be equally capable of producing a new view resembling an internal or external view, and thus both internal and external facilitation should occur. The present finding of only internal facilitation is an evidence against this theory. If the instability of computations for the external views is considered, his theory will be similar to the active interpolation account. With the account one should expect uniform performance for all the studied views and views in between during the test phase for non-occluded rotation but not occluded rotation. However, perfect internal facilitation was found for the amoeboid objects with occluded rotation in Experiment 1, which is difficult to explain with this account. Another possible explanation is a version of the traditional prototype account, which suggests that studying two views might cause the visual system to develop sensitivity not to the two views but to the their average. According to this account, since the internal view usually resembles the average view, recognition performance should be the best with the internal view, while the external view does not benefit from studying two views. Better performance for the internal view than the studied view should be expected, but this only occurred in some conditions in the present study but not others.

Link to Existing Theories

As mentioned before, the pooled activation account is contributed by ideas from different researchers. Both Poggio and Edelman (1990) and Perrett et al. (1998) provided detailed description of the account, although the former presented in computational terms while the latter in neurophysiological terms. Actually the two theories can be regarded as very similar, with Poggio and Edelman (1990) describing at the computational and algorithmic level while Perrett et al. (1998) at the neural or implementional level. As described in Edelman and Poggio (1990), object recognition at a novel view can be performed by a RBF network involving two stages. At the first stage, individual view-tuned units, or nonlinear receptive fields, have intermediate responses determined by the difference between the novel view and the preferred view of the each unit. These individual responses are then gathered by weighted summation to generate the output. Recognition performance in terms of both accuracy and reaction time are predicted to be higher for internal views since they are likely to cause more intermediate responses, leading to larger overall output, although in simulation (e.g., Bricolo et al., 1997; Poggio & Edelman, 1990) and human behavioral studies (e.g., Bülthoff & Edelman, 1992) only accuracy was concerned. Similarly, Perrett et al. (1998) hypothesized that populations of cells are present to code for an object. Each cell population is broadly tuned to a particular object view, thus it can be regarded as equivalent to an individual view-tuned unit or nonlinear receptive field in an RBF network. A novel view causes different levels of activity in different cell populations, depending also on the angle between the novel view and the preferred view of the population. The different activity level is defined as the firing rate of cells. As long as different views of an object have been associated to the same response, cells in the higher level sum up activity of individual cell populations over time, and the time for this summed activity to reach a threshold level determines the recognition time. This process is very similar to the second stage described in Edelman and Poggio (1990) in which individual receptive field activity is combined linearly. Views between the studied views, according to Perrett et al.'s (1998) theory, should cause more activations of individual neuronal populations, leading to shorter time for the convergent output to reach the threshold level, and thus faster reaction time. Although no explicit prediction about recognition accuracy has been discussed, it is plausible that greater convergent activity is more distinguishable from the noise or background neuronal activity and should contribute to fewer errors. Internal facilitation found in the present study in terms of both accuracy and reaction time can be explained by both theories. So can the similar recognition performance with one studied view or two close studied views. Poggio and Edelman (1990) allowed the possibility that the number of view-tuned units formed can be smaller than the number of trained views, and that the centres of the radial basis function of a receptive field can be updated by untrained views similar to the preferred view of that receptive field. Perrett et al. (1998) also mentioned that individual neuronal populations can be modified by experience.

The present study provided additional evidence against Biederman's RBC model. Viewpoint dependent recognition was shown in the geon objects used in Experiments 3 and 4. When only one view had been studied, recognition performance deteriorated systematically with occluded and non-occluded rotation. It should be noted that all the views of each geon object in the non-occluded rotation set revealed the same geon structural descriptions (GSDs) defined by Biederman and Gerhardstein (1993). Also, individual objects differed in non-accidental properties defined by Biederman and Bar (1999). According to the RBC model, participants should have accomplished view-invariant discrimination at the first time they saw the objects. However, large viewpoint costs were involved for rotated views (rotation rates did not exceed 200°/s for geon objects with non-occluded rotations in Experiments 3 and 4). The internal facilitation found also helps defend the view-based class of theories from criticisms about the lack of identified mechanisms to recover viewpoint costs. The pooled activation account can act as a plausible alternative to the less biologically plausible hypotheses like mental rotation and linear combination.

Further Study

A number of processes have been described to account for the interpolation found. They include, for example, generation of new view-specific representations, association of two or more view representations, and activation of both representations and their common afferent unit by the internal view. Such processes may likely be employed with different stimuli. For example, when one view has been studied, a second studied view has to be different enough from the first for a new view representation to be formed. Also, with a larger difference between the two studied views, the difficulty to link the two corresponding representations also increases. Although such processes probably involve the same brain areas, they may occur at different times after stimuli presentation. Electrophysiological recording, which is sensitive to temporal changes in brain activities, may therefore help to identify brain activities associated with the above-mentioned processes.

The present study employed a sequential matching task, which taxed participants' working memory. Nevertheless, long-term object memory was addressed in the neural research and many of the behavioral studies (e.g., Kourtzi & Shiffrar; Schwoebel & Srinivas, 2000; Srinivas & Schwoebel, 1998) reviewed in the Introduction. It is assumed that the short-term and long-term object recognition system employed the same neural resources and processes in view interpolation.

Similar results are expected for long-term recognition tasks as these tasks revealed similar viewpoint dependent response pattern with single studied views (Hayward & Tarr, 1997; Tarr et al., 1998). These tasks are also useful in that the association strength of the studied views can be manipulated by, for example, varying the temporal contiguity between the studied views as Schwoebel and Srinivas (2000) did. Also, it would be easier to investigate the effect of studying more than two views with long-term memory tasks than short-term tasks.

To conclude, the present study provides stronger support for the pooled activation account by revealing interpolation in amoeboid and geon objects with and without self-occlusion across depth rotation. It is also the first study to compare the extent of interpolation between objects with occluded and non-occluded rotations. The findings were compatible with the biologically plausible hypothesis that a novel view between familiar views activates different view-specific representations to different degrees and the pooled activation results in recognition to be accomplished.

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39,* 2885-2899.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 1506-1514.

Bricolo, E., Poggio, T., & Logothetis, N. (1997). 3-D object recognition: a model of view-tuned neurons. In Mozer, M. C., Jordan, M. I., & Petsche, T. (Eds.), *Advances in Neural Information Processing Systems 9*, pp. 41-47. MA, Cambridge: MIT Press.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89,* 60-64.

Edelman, S. (1999). *Representation and recognition in vision.* London:MIT.

Edelman,S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385-2400.

Fujita, I, Tanaka, K, Ito, M., & Cheng , K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360,* 343-346.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception & Performance, 23,* 1511-1521.

Kourtzi, Z., & Shiffrar, M. (1999). The visual representation of three-dimensional, rotating objects. *Acta Psychologica, 102,* 265-292.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica, 102,* 221-245.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: evidence for viewpoint-specific recognition procedures. *Perception, 23,* 595-614.

Liu, Z. (1996). Viewpoint dependency in object representation and recognition. *Spatial Vision*, 9, 491-521.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 3,* 270-288.

Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception, 26,* 1231-1257.

Oram, M. W., Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68,* 70-84.

Perrett, D. L., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalization of recognition without mental transformations. *Cognition, 67,* 111-145.

Perrett, D. L., Oram, M. W., Harries, M. H., Beyan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centered and object centered coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86,* 159-173.

Poggio, Y., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263-266.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019-1025.

Srinivas, K, & Schwoebel, J. (1998). Generalization to novel views from view combination. *Memory and Cognition, 26,* 768-779.

Schwoebel, J., & Srinivas, K. (2000). Recognizing objects seen from novel viewpoints: Effects of view similarity and time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 915-928.

Tarr., M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology, Human Perception and Performance, 21,* 1494-1505.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233-282.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1,* 275-277.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, MA: MIT Press.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67,* 21-44.

Wang, G, Tanaka, K, & Tanifuji, M. (1994). Optical imaging of functional organization in monkey inferotemporal cortex. *Science, 272,* 1665-1668.

Weiskrantz, L., & Saunders, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain, 12,* 299-311.

Williams, P., & Tarr, M. J. (1999). *RSVP: Experimental control software for*

MacOS [Online]. Available: http://psych.umb.edu/rsvp/

Table 1.

<u>The Conditions in Experiments 1 to 4.</u>

Study condition		**Test view**				
		View 1	*View 2*	*View 3*	*View 4*	*View 5*
Experiment 1 (amoeboid)	*Single*	External	Studied	Internal		
				or		
				Internal	Studied	External
	Double *(40° apart)*	External	Studied	Internal	Studied	External
Experiment 2 (amoeboid)	*Single*		Studied	Novel		
				or		
				Novel	Studied	
	Double-Adjacent *(20° apart)*	Studied (Far)	Studied (Near)	Novel **(External)**		
				or		
				Novel **(External)**	Studied (Near)	Studied (Far)
	Double-Separated *(60° apart)*		Studied (Near)	Novel **(Internal)**		Studied (Far)
				or		
		Studied (Far)		Novel **(Internal)**	Studied (Near)	
Experiment 3 (geon)	*Single*	External	Studied	Internal		
				or		
				Internal	Studied	External
	Double *(56° apart)*	External	Studied	Internal	Studied	External
Experiment 4 (geon)	*Single*		Studied	Novel		
	Double-Adjacent *(28° apart)*	Studied (Far)	Studied (Near)	Novel **(External)**		
				or		
				Novel **(External)**	Studied (Near)	Studied (Far)
	Double-Separated *(84° apart)*		Studied (Near)	Novel **(Internal)**		Studied (Far)
				or		
		Studied (Far)		Novel **(Internal)**	Studied (Near)	

Figure Captions

Figure 1. Bülthoff and Edelman's (1992) study. (a). Examples of amoeboid and wire-like objects used. Some parts are occluded at some views. Object is located at the center of the sphere; parts on the surface of the sphere are conceived as possible views of object. (b). View sphere. The filled circles represent the training views, the circles the internal views, the squares the external views, and the triangles the orthogonal views. Adapted from Bülthoff and Edelman (1992).

Figure 2. The general presentation sequence in all the experiments. For single study condition, the same study view was presented two times in the study phase. For double study condition, different views were presented in the study phase. Participants had to provide same/different response on seeing the test stimulus.

Figure 3. (a). Examples of the amoeboid objects used in Experiments 1 and 2. Different protrusions were similar in general, so the main difference between objects is in the configuration. For objects in the occluded rotation set, the protrusions marked by squares and circles were clearly visible at some of the views but heavily or totally occluded at other views. Adjacent views were 20° apart. (b). Examples of the geon objects used in Experiments 3 and 4. Objects can be discriminated from each other by the different geon protrusions in different locations. The central part was informative of the object view. For objects in the occluded rotation set, the protrusions marked by squares and circles were clearly visible at some of the views but heavily or totally occluded at other views. Adjacent views were 28° apart. Note that self-occlusion was more serious in geon than in amoeboid objects, in that two protrusions were usually occluded in a geon object while only one was occluded in an amoeboid object.

Figure 4. Results of Experiment 1, using amoeboid objects. View 2 or view 4 was

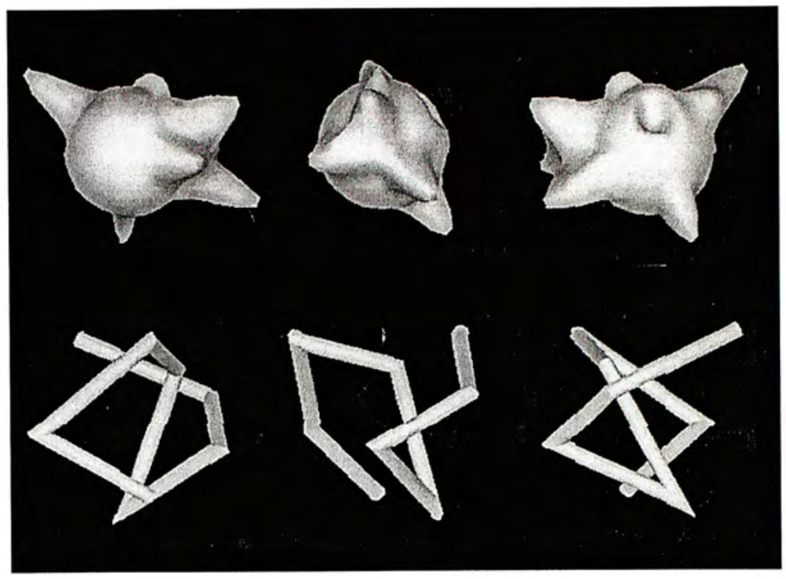
studied in the single study condition, and both views 2 and 4 (40° apart) were studied in the double study condition. Error bars denote the standard error of the mean. (a). For objects in the non-occluded rotation set, performance was worse at the external or internal views than at the studied view when only one view was studied. With two studied views, recognition at internal view became as good as that at the studied view, while recognition at the external view remained the worst. (b). Similar results were found for objects with occluded rotation.

Figure 5. Results of Experiment 2, using amoeboid objects. In the double-adjacent study condition the studied views were 20° apart and in the double-separated condition they were 60° apart. (a). In general, for objects with non-occluded rotation, any viewpoint cost between the nearest studied view and the novel view disappeared only when the two studied views were far apart and embedded the novel view, an indication of internal facilitation without external facilitation. Note that whether only one view or two close views were studied made little difference in subsequent recognition performance. (b). For objects with occluded rotation, neither internal nor external facilitation was present. Performance at the novel view was in general worse than the nearest studied view in all study conditions.

Figure 6. Results of Experiment 3, using geon objects. View 2 or view 4 was studied in the single study condition, and both views 2 and 4 (56° apart) were studied in the double study condition. (a). For objects with non-occluded rotation the worse performance at the external views than at the studied view in the single study condition did not change when two views were studied. Internal facilitation was shown by the improved performance at the internal view with two studied views. (b). For objects with occluded rotation only weak internal facilitation was found in RT.

Figure 7. Results of Experiment 4, using geon objects. In the double-adjacent study condition the studied views were separated by 28° and in the double-separated condition they were separated by 84°. (a). For objects with non-occluded rotation, generalization of recognition to the novel view was the best when the studied views were far apart. Response at the novel view was worse than at the nearest studied view with only one view studied or with two close views studied. (b). For objects with occluded rotation, neither internal nor external facilitation was present. Performance at the novel view was in general worse than at the nearest studied view in all study conditions.

Figures 1a and 1b



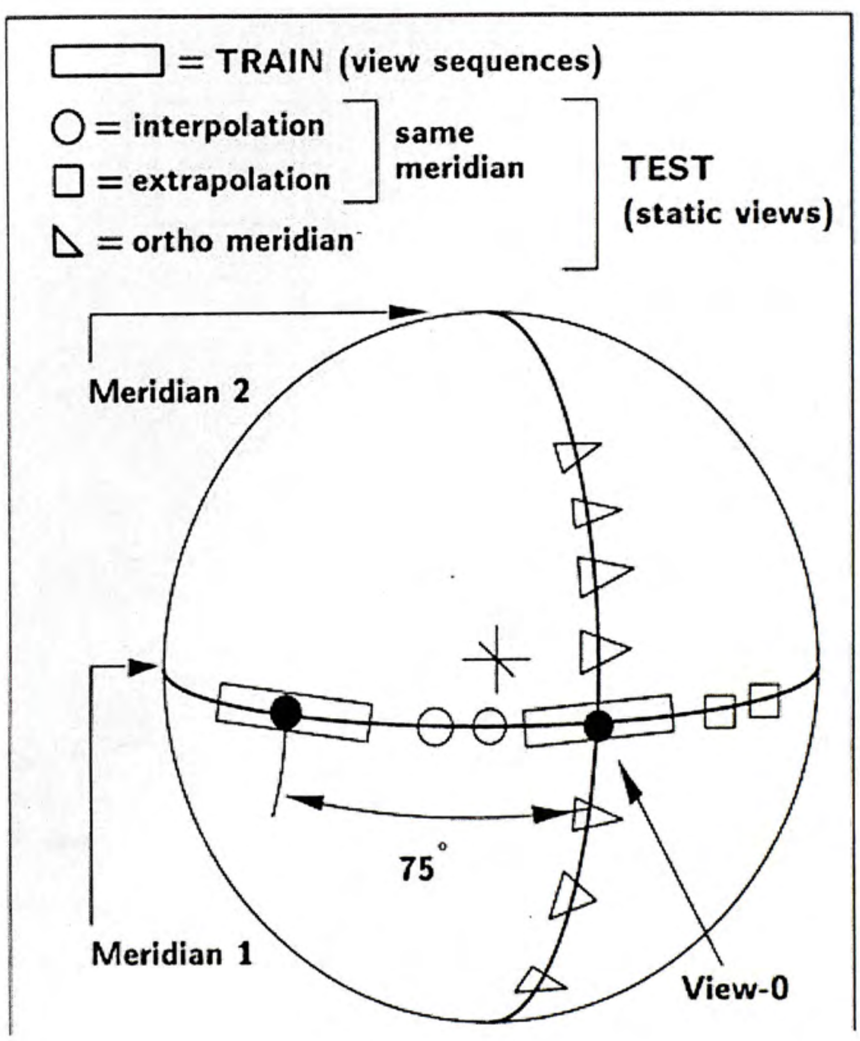


Figure 2

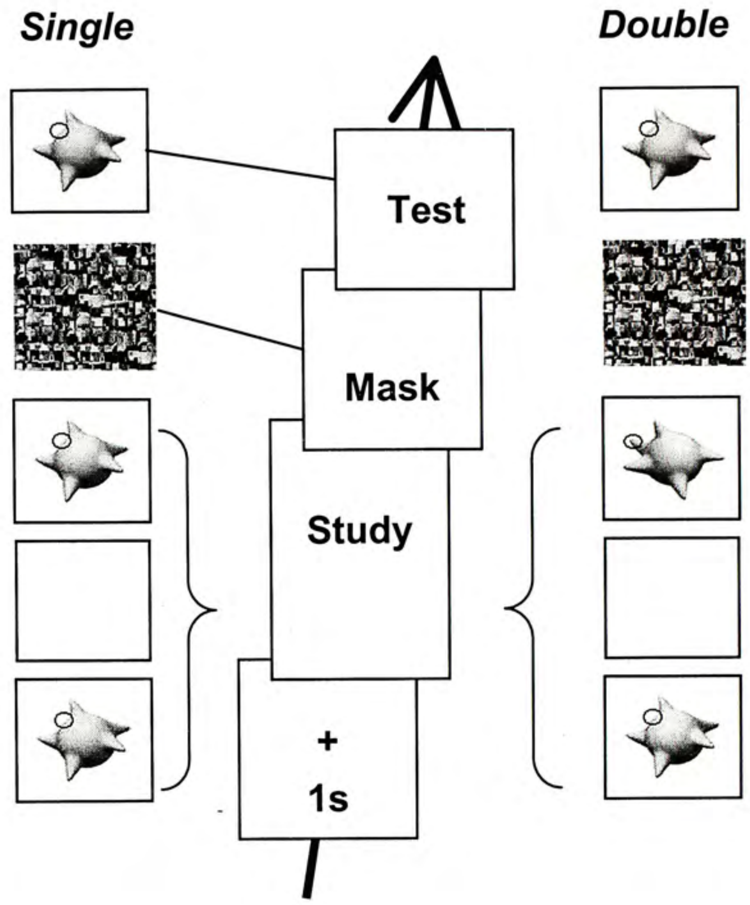
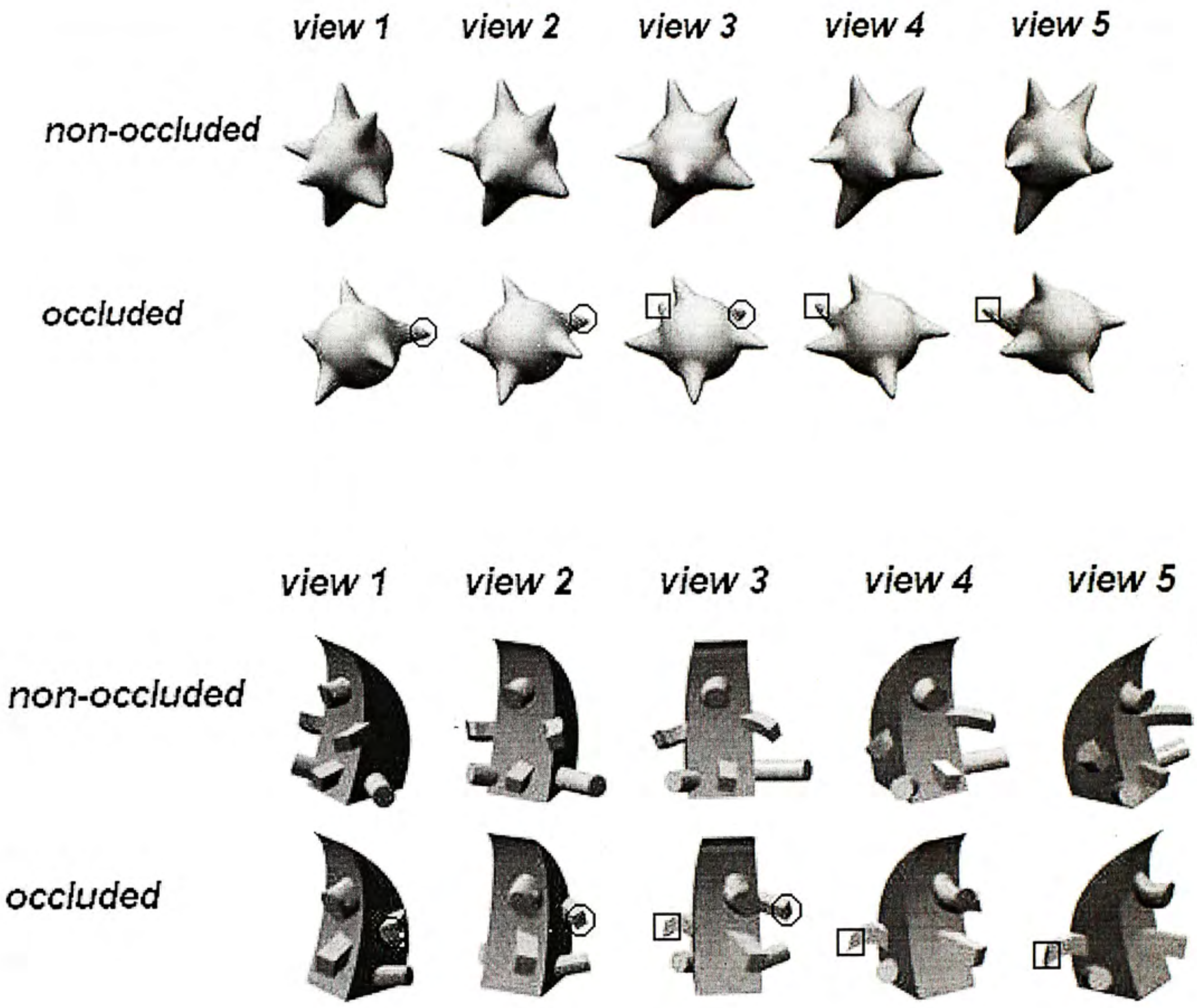
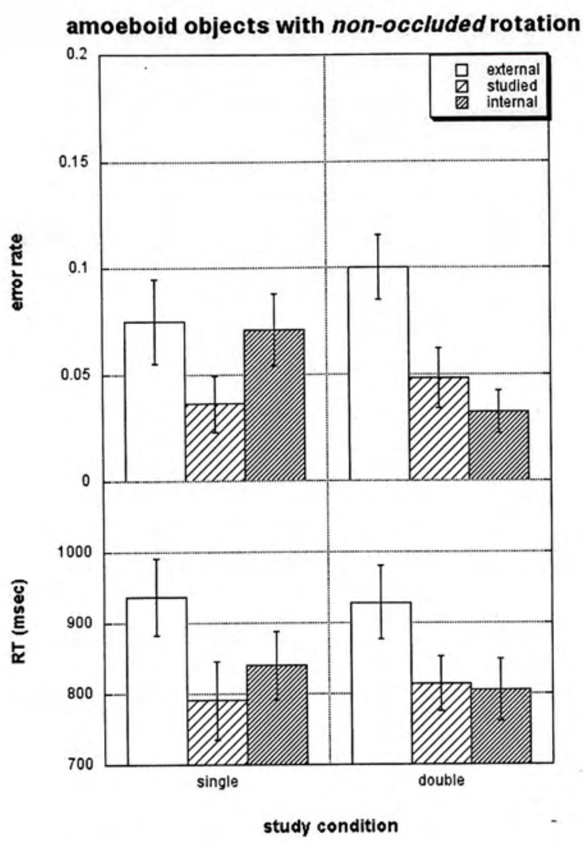
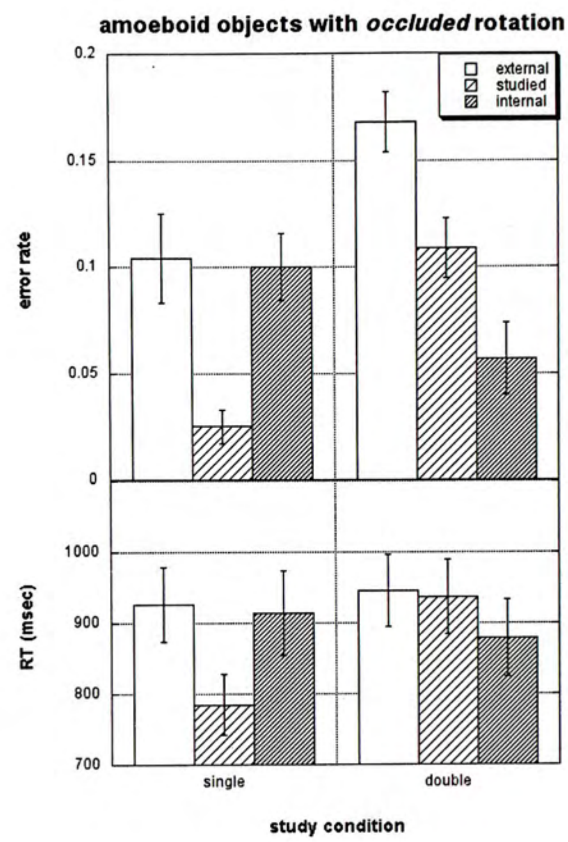


Figure 3

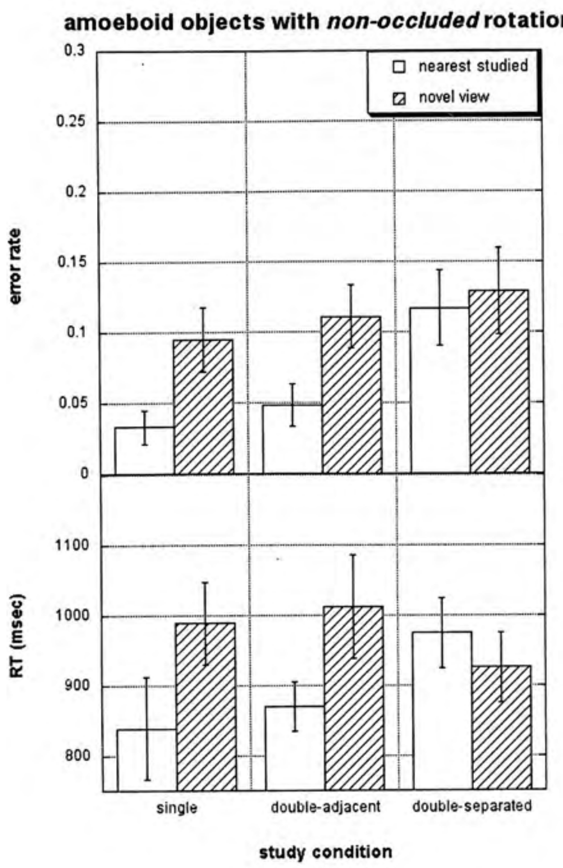

view 1
view 2
view 3
view 4
view 5
non-occluded
occluded
view 1
view 2
view 3
view 4
view 5
non-occluded
occluded

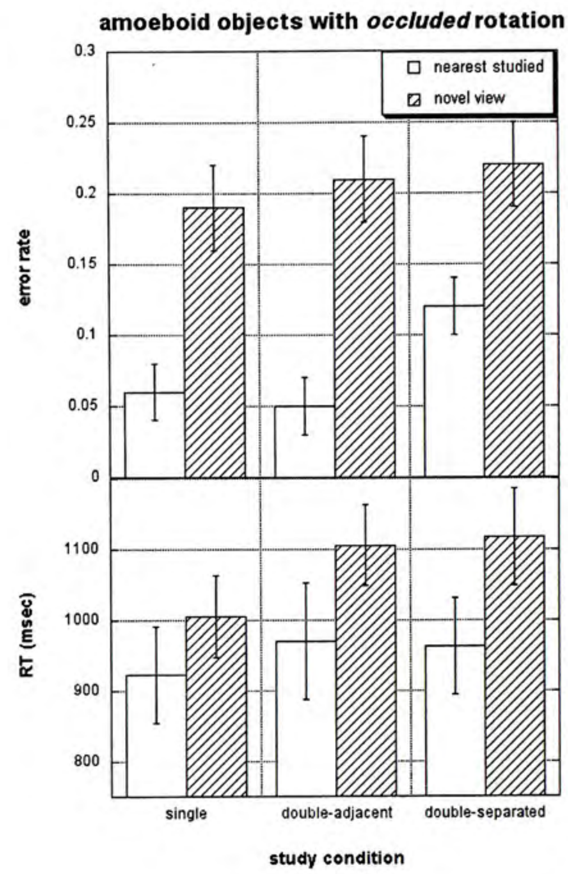
Figures 4a and 4b



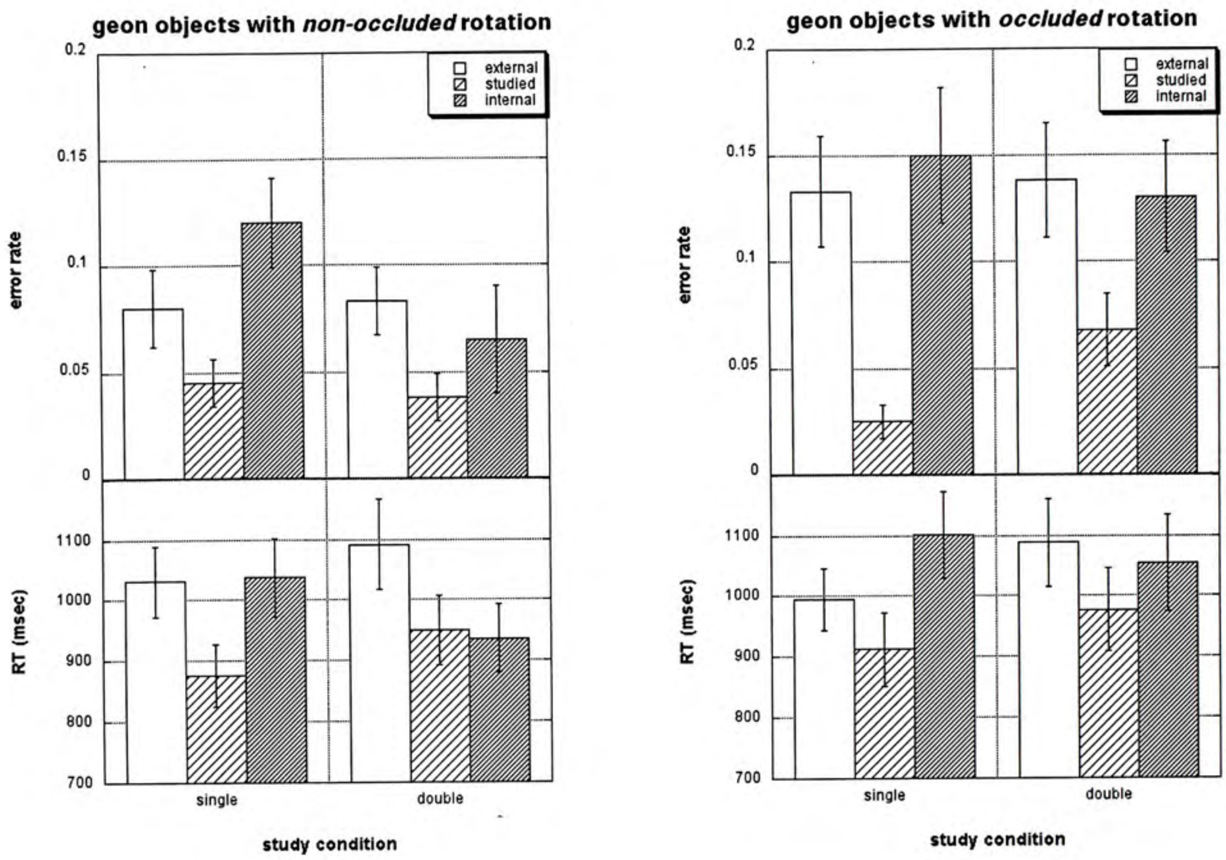


Figures 5a and 5b





Figures 6a and 6b



Figures 7a and 7b